ESTIMATION OF FUNCTION (F)

→ Function f, connects the inputs variable to the output variable.
→ f may involve more than one input variable.

Why estimate f?

Two main reasons to estimate f : i) Prediction ii) Inference.

i) Prediction –

- In many situations, a set of input X are readily available, but the output Y cannot be easily obtained. In this setting, since the error term average to zero, we can predict Y using $\boxed{\hat{Y} = \hat{f}(x)}$

- Example, suppose $X_1, X_2, X_3 \dots X_n$ are characteristics of a patient blood sample and Y is a variable encoding the patient risk for adverse reaction to a particular drug. It is natural to seek predict Y using X, since we can then avoid giving the drug in question to patient who are at high risk of adverse reaction.

- Accuracy of $\hat{Y}$ as a prediction depend on two quantities i) reducible error
ii) irreducible error.

→ Irreducible error is something no matter how well we estimate f, we cannot reduce the error. The quantity may also contain unmeasurable variation. For example, the risk of an adverse reaction might vary from patients on a given day, depending on manufacturing variation in the drug itself or the patient's general feeling of well-being of the day.

$$\boxed{E(Y-\hat{Y})^2 = E|f(x)+\varepsilon-\hat{f}(x)|^2 = \underbrace{|f(x)-\hat{f}(x)|^2}_{\text{Reducible error}} + \underbrace{Var(\varepsilon)}_{\text{Irreducible error}}}$$

where $E(Y-\hat{Y})^2$ represent average or expected value of squared difference between predicted and actual value of Y and $var(\varepsilon)$ represent variance associate with error term ε.

ii) Inference –

- Often interested in understanding the way that Y is affected as $X_1, \dots, X_p$ changes.
Few questions like –
i) Which predictors are associated with response?
ii) What is the relationship between response and each predictor? (Correlation)
iii) Can the relationship between Y and each predictor be adquently summarized using a linear equation or relationship is more calculated?

How do we estimate f?

→ Our goal is to apply a statistical learning method to the training data in order to estimate the unknown function f. In other word, we want to find a function $\hat{f}$ such that $Y = \hat{f}(x)$ for any observation (X, Y).

→ Broadly this can be classified into two characteristics - i) Parametric methods
ii) Non parametric methods.

i) Parametric methods → Two step model base approach.

Step 1 - Make an assumption about functional form or shape of f. For example, f is a linear function of x. $f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_N$.

Step 2 - After a model has been selected, we need a procedure that uses the training data to fit or train the model. In this case of linear model, we need to find /estimate the parameter $\beta_0, \beta_1, \dots, \beta_n$.

For example, Income (of a person) $\approx \beta_0 + (\beta_1 \times \text{education}) + (\beta_2 \times \text{experience})$

It is the linear relationship between response and two predictors.

ii) Non parametric methods.

→ Do not make any explicit assumption about the function form of f.

→ Such approach have major advantage over parametric appraches: by avoiding the assumption of a particular form of f. Any parametric approach bring use to estimate f very different from true f, in this case model will not fit the data well. In this non parametric avoid the situation because it does not have any assumption

→ But the biggest disadvantage of non parametric is they do not reduce the problem of estimating f to smaller number of observations/parameters. A very large number of observations (far more than parametric approach) is required in order to obtain an accurate estimate of f.